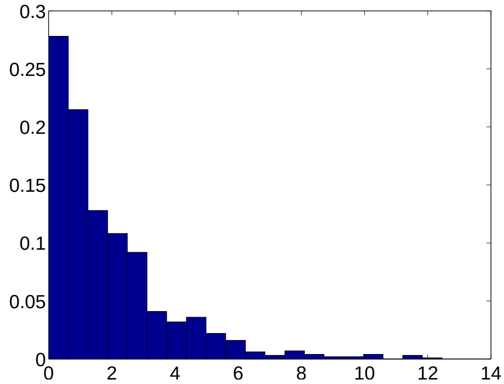
STAT 391
Solutions to the Final Exam
June 7, 2007

©Marina Meilă and Debashis Mondal
mmp@cs.washington.edu

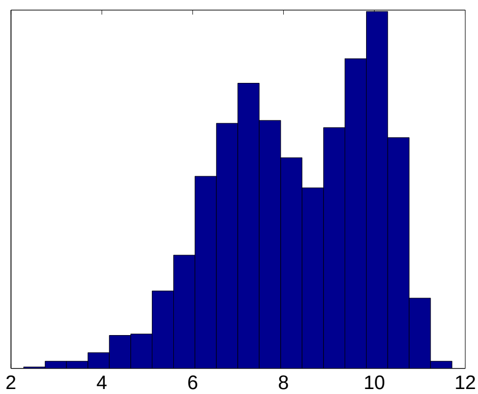
You are allowed 10 pages of notes. Write your name clearly on of the notes pages; you will be asked to hand them in with the exam, but they will not be graded and will be returned to you in the statistics office after 4pm the day of the exam.
No electronic devices of any kind are allowed during the exam.
Any fact that was proved in the lectures or in the notes can be used without proof.

Problem 1

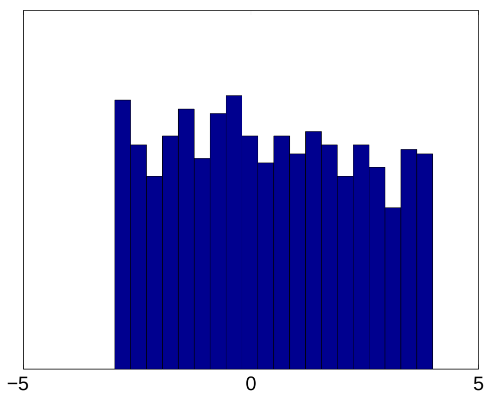
The following plots represent histograms of densities on the real axis. Mark the most appropriate answer in each case.



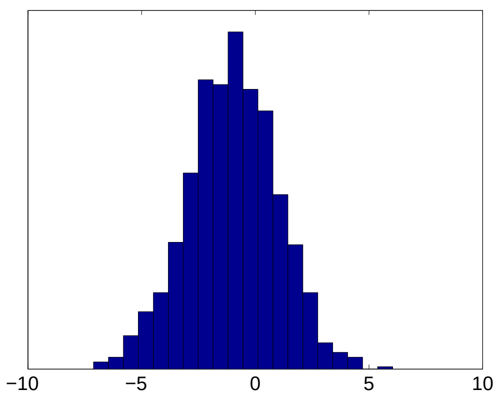
This density is
exponential $\lambda e^{-\lambda x}$ $\quad \lambda =$ 0.5



This density is
other



This density is
uniform $\quad a =$ -3 $\quad b =$ 4



This density is
normal $\quad \mu =$ -1 $\quad \sigma =$ 2

Problem 2

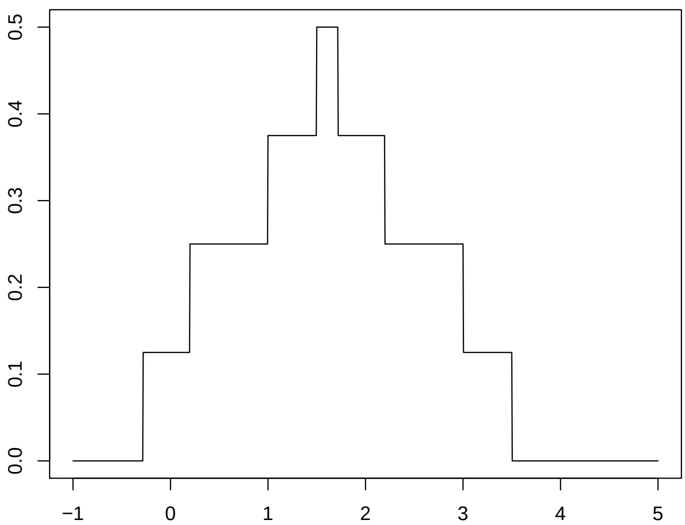
The figure depicts a data set of $n = 4$ points on the real axis. Draw a kernel density estimate $f(x)$ based on this data set, using a **square kernel** with $h = 2$.

You can choose to draw an unnormalized density; in this case write value of the normalization constant Z above the graph. If your density is normalized, write $Z = 1$ above the graph.

The square kernel is given by
$k(z) = \begin{cases} 1 & \text{if } z \in [-0.5, 0.5] \\ 0 & \text{otherwise} \end{cases}$

$$f(x) \;=\; \frac{1}{4 \times 2} \sum_{i=1}^{4} 1_{[|x - x_i|/2 \leq 0.5]}$$

$$Z \;=\; 1$$



Problem 3.

A coin is tossed $n = 9$ times and the sequence of outcomes is $X_{1:n} = (1, 0, 1, 1, 1, 0, 1, 1, 0)$.

3.1 Estimate $p = P(1)$, the probability of outcome 1.

$$\hat{p} \;=\; \frac{n_1}{n} \;=\; \frac{6}{9} \;=\; \frac{2}{3}$$

Use your estimate $\hat{p}$ to answer the following questions:

3.2 What is the probability of obtaining the sequence $X'_{1:4} = (1, 0, 0, 1)$.

$$P(X'_{1:4}) = P(X_1)P(X_2)P(X_3)P(X_4) = (1-p)^2 p^2 = \frac{2 \times 1 \times 1 \times 2}{3^4} = \frac{4}{81}$$

3.3 What is the probability of obtaining a sequence containing 6 ones and 3 zeros?

Only the symbolic answer is needed.

This is the probability of the event $n_0 = 3$, $n_1 = 6$ and can be computed from the Binomial distribution

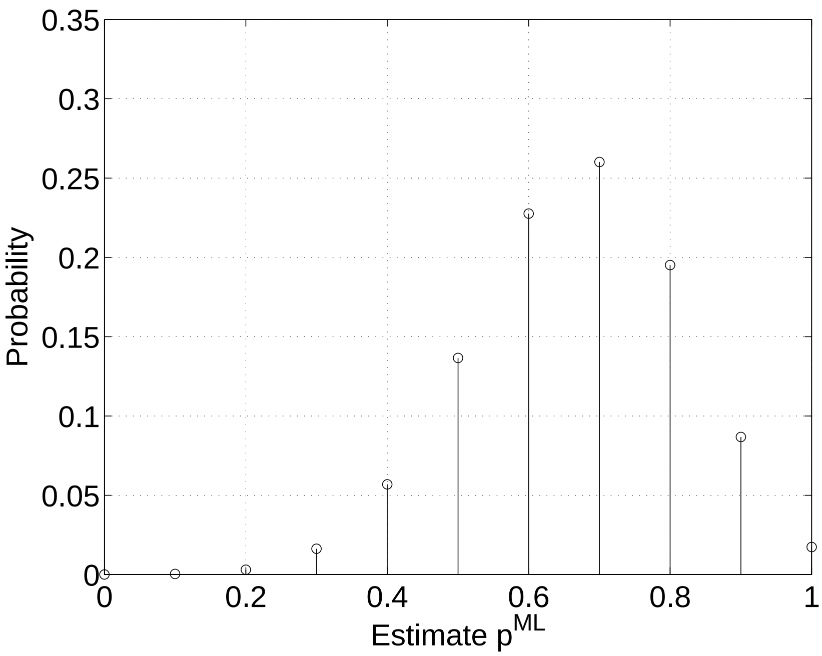
$$P[n_0 = 3, n_1 = 6] = \binom{9}{3\,6} (1-p)^3 p^6$$

3.4 What is the probability of obtaining a sequence of 6 outcomes that starts and ends with a 1?

Because the coint tosses are independent, this probability is

$$P(X_1 = 1)P(X_2 \in \{0,1\})P(X_3 \in \{0,1\}) \ldots P(X_5 \in \{0,1\})P(X_6 = 1) = p^2 = 4/9$$

3.5 You assume that the estimate of p that you found is the true one, and want to obtain confidence intervals. For this you make the plot below, giving the distribution of p^{ML} for $n = 10$ with your estimate $\hat{p}$ as the true value. (I made this distribution assuming that you found the correct value in question 3.1).



Use this plot to answer the following questions.

Mark the value $\hat{p}$ estimated in 3.1 on the graph. What is the (approximate) probability that $|p^{ML} - \hat{p}| < 0.1$?

This probability is $P[|p^{ML} - \hat{p}| < 0.1] = P(p^{ML} = 0.6) + P(p^{ML} = 0.7) \simeq 0.49$.

3.6 Are the following intervals 95% confidence intervals for p? Circle the appropriate answer.

Interval	
[0.55, 1]	Surely NO
[0.1, 0.9]	Surely YES
[0.4, 0.9]	Surely YES / OR /Could be, but it's too close to tell

Problem 4

Let $S = (-\infty, 0]$ and let $X_1, \ldots X_n$ be drawn independently from

$$f_X(x) \;=\; e^x \text{ for } x \in S$$

4.1 Denote by F_X the CDF of X. What is the expression of F_X?

For $x \in S$

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(u)du \;=\; \int_{-\infty}^{x} e^u du \\ &= e^u|_{-\infty}^{x} \;=\; e^x \end{aligned}$$

So

$$F_X(x) \;=\; \begin{cases} e^x, & x \le 0 \\ 1 & x > 0 \end{cases}$$

4.2 Denote by Z the maximum of $X_1, \ldots X_n$. What is the probability that $Z \le -1$?

Since $X_1, \ldots X_n$ are i.i.d

$$\begin{aligned} P(Z \le -1) &= P[(X_1 \le -1) \text{ AND } (X_2 \le -1) \text{ AND } \ldots (X_n \le -1)] \\ &= P[X_1 \le -1] P[X_2 \le -1] \ldots P[X_n \le -1] \\ &= [F_X(-1)]^n \;=\; e^{-n} \end{aligned}$$

4.3 What is the probability that $Z \le a$, for $a \in (-\infty, 0]$? Use this result to derive F_Z the CDF of Z.

Similarly, $P[Z \le a] = F_X(a)^n = e^{an}$ for $a \le 0$ and $P[Z \le a] = 1$ for $a > 0$. Therefore, the CDF of Z is

So

$$F_Z(x) \;=\; \begin{cases} e^{na}, & a \le 0 \\ 1 & a > 0 \end{cases}$$

4.4 What is the probability density function of Z (denote it by f_Z)?

$$f_Z(a) \;=\; \frac{d}{da}F_Z(a) \;=\; \frac{d}{da}e^{na} \;=\; ne^{na} \quad \text{for } a \leq 0.$$

4.5. What is the median of Z ?

The median of Z is the point for which $F_Z(a) = 1/2$ i.e

$$e^{nm_z} \;=\; 1/2 \;\Rightarrow\; nm_z \;=\; -\ln 2 \;\Rightarrow\; m_z \;=\; -\frac{\ln 2}{n}.$$

4.6. Denote by A the event "at least one of $X_{1:n}$ is between -1 and 0". What is the smallest n so that $P(A) \geq 0.5$?

A implies that $Z \in [-1, 0]$. Now $Z \geq -1$ w.p. at least 1/2 means that the median of Z is at least -1. (Or, equivalently, $F_Z(-1) \leq 0.5$.). Therefore

$$-\frac{\ln 2}{n} \geq -1$$

or equivalently

$$e^{-n} \leq 0.5$$

The smallest n for each these conditions are satisfied is $n = 1$ because $\ln 2 < 1$ and $e > 2$.

Problem 5

5.1 Let X be a random variable with $E[X] = 3$, $Var(X) = 4$. Compute the expectation and variance of $Y = 2 - 3X$.

$$E[Y] \;=\; E[2-3X] \;=\; 2 - 3E[X] \;=\; -7.$$

$$Var(Y) \;=\; Var(2-3X) \;=\; Var(-3X) \;=\; (-3)^2 Var(X) \;=\; 36.$$

5.2. $Z \sim Normal(-2, 1)$. What is the probability that $Z \leq 0$? Use the table provided.

We find the transformation that makes Z into a $Normal(0, 1)$ random variable. This is $\tilde{Z} = \frac{Z-(-2)}{1} = Z + 2$. In this simple case, the transformation is equivalent to shifting Z into the origin.

$$P[Z \leq 0] \;=\; P[Z+2 \leq 2] \;=\; P[\tilde{Z} \leq 2] \;=\; 1-P[\tilde{Z} > 2] \;=\; 1-P[\tilde{Z} < -2] \;=\; 1-\text{table}(-2) \;=\; 0.9772$$

5.3. $Z \sim Normal(4, 4)$. What is the probability that $Z \geq 1$? Use the table provided.

We find the transformation that makes Z into a $Normal(0,1)$ random variable. This is $\tilde{Z} = \frac{Z-(4)}{2}$ because $\sigma^2 = 4 \Rightarrow \sigma = 2$.

$P[Z \geq 1] \;=\; P[Z/2-2 \geq 1/2-2] \;=\; P[\tilde{Z} \geq -3/2] \;=\; 1-P[\tilde{Z} \leq -3/2] \;=\; 1-\text{table}(-1.5) \;=\; 0.9332$

The CDF of the standard normal $X \sim Normal(\mu = 0, \sigma^2 = 1)$

a	0	-0.5	-1.0	-1.5	-2.0	-2.5	-3.0	-3.5	-4.0
$P[X \leq a]$	0.5000	0.3085	0.1587	0.0668	0.0228	0.0062	0.0013	0.0002	0.0000

Problem 6

Indiana Jones is probing an ancient castle for treasures. He has two detectors which are supposed to beep if there is a treasure in the wall. The Gold Detector will alwasy beep if a treasure is present, but will beep with probability (w.p.) 1/3 even if no treasure is there. The Hole Detector will beep w.p. 3/4 if there is treasure and will beep w.p 1/4 if there is no treasure. (You may recall these probabilities are called *detection rate* and *false alarm rate*.) The probability of treasure in any given spot is 1/10 and we assume that the former owners of this castle didn't leave any hole in the wall unfilled with gold. Denote by T, G and H the events "there is treasure", "the Gold Detector beeps", and "the Hole Detector beeps" respectively. If there is no treasure, the two detectors are right or wrong idenpendently of each other.

Recommended but not required: make a drawing of the sample space here.

	GH	$\bar{G}H$	$\bar{G}\bar{H}$	$G\bar{H}$
T		$-$	$-$	
$\bar{T}$				

There are 6 possible outcomes (the outcomes marked with – are impossible).

6.1. Compute the probability that the Gold Detector is correct.

$$\begin{aligned} P(G\,\text{correct}) &= P(TG) + P(\bar{T}\bar{G}) \\ &= P(T)P(G|T) + P(\bar{T})P(\bar{G}|\bar{T}) \;=\; 0.1 \times 1 + 0.9 \times \frac{2}{3} \;=\; 0.7 \end{aligned}$$

6.2. Compute the probability that both detectors are beeping.

$$\begin{aligned} P(GH) &= P(TGH) + P(\bar{T}GH) \\ &= P(T)P(G|T)P(H|T) + P(\bar{T})P(G|\bar{T})P(H|\bar{T}) \\ &= 0.1 \times 1 \times \frac{3}{4} + 0.9 \times \frac{1}{3}\frac{1}{4} \;=\; 0.15. \end{aligned}$$

6.3 Both detectors beep. What is probability that Indiana Jones has found treasure?

By Bayes' rule

$$P(T|GH) \;=\; \frac{P(T)P(GH|T)}{P(GH)} \;=\; \frac{P(T)P(H|T)P(H|T)}{P(GH)} \;=\; \frac{0.1 \times 1 \times \frac{3}{4}}{0.15} \;=\; \frac{1}{2}.$$

6.4 Indiana Jones is in front of a narrow passage. He will have to take only one detector with him. Which detector should he take? Mr. Jones would like to maximize his expected gain and he knows that: (1) he will dig if the detector beeps and will not dig otherwise; (2) digging in any place will cost him \$300 (3) treasures in this kind of castle are worth \$1,000; (4) the other costs of exploring beyond the passage will be the same no matter what the outcome.

There were two possible interpretations of this question and both were considered OK. One is the gain in the outcome TG is \$700 (this should have been correct) the other one is that the gain for TG is \$1000.

<u>If he takes only the Gold Detector:</u>
Outcome space with gain for each outcome:

TG $70\ or\ 1000$	$\bar{T}G$ -300
	$\bar{T}\bar{G}$ 0

$$E[\text{gain}] \;=\; 700P(TG) - 300P(\bar{T}G) + 0 \;=\; 700 \times 0.1 - 300 \times 0.9 \times \frac{1}{3} + 0 \;=\; -20$$

OR

$$E[\text{gain}] \;=\; 1000P(TG) - 300P(\bar{T}G) + 0 \;=\; 1000 \times 0.1 - 300 \times 0.9 \times \frac{1}{3} + 0 \;=\; 10.$$

<u>If he takes only the Hole Detector:</u>
Outcome space with gain for each outcome

TH $700\ or\ 1000$	$\bar{T}H$ -300
$T\bar{H}$ 0	$\bar{T}\bar{H}$ 0

$$E[\text{gain}] \;=\; 700P(TH) - 300P(\bar{T}H) + 0 \;=\; 700 \times 0.1 \times \frac{3}{4} - 300 \times 0.9 \times \frac{1}{4} + 0 \;=\; -15 > -20$$

OR

$$E[\text{gain}] \;=\; 1000P(TH) - 300P(\bar{T}H) + 0 \;=\; 1000 \times 0.1 \times \frac{3}{4} - 300 \times 0.9 \times \frac{1}{4} + 0 \;=\; 7.25 < 10.$$

For gain(TG)=700: Taking the Hole Detector gives him a higher expected gain so he should chose it.

For gain(TG)=1000: Taking the Gold Detector gives him a higher expected gain so he should chose it.

Problem 7 (after Al Drake)

A cookie factory makes raisin cookies. The number n of raisins in a cookie is independent of the number of raisins in any other cookie and follows a Poisson distribution with parameter $\lambda = 3$. What is the probability that a given raisin is in a cookie containing 3 raisins?

FYI: The Poisson distribution has $S = \{0, 1, 2, \ldots\}$ and is defined by

$$P(n) \; = \; e^{-\lambda} \frac{\lambda^n}{n!}$$

The expectation and variance of the Poisson distribution are $E[n] = \lambda,\; Var(n) = \lambda$.

Let's list the space of outcomes for a raisin. A rasin can be:

- single in a 1 raisin cookie
- (with equal probability) 1st or 2nd in a 2 raisin cookie
- (with equal probability) 1st, 2nd or 3rd in a 3 raisin cookie
- etc.

Hence

$$P[\text{raisin in cookie with 3 raisins}] \; = \; \frac{3P(3)}{0P(0) + 1P(1) + 2P(2) + \ldots + nP(n) + \ldots}$$
$$= \; \frac{3P(3)}{E[n]} \; = \; 3e^{-3}3^3/3!/3 \; = \; 4.5 \times e^{-3}$$